Health and Environment Linked for Information Exchange (HELIX)-Atlanta: A CDC-NASA Joint Environmental Public Health Tracking Collaborative Project

Mohammad Al-Hamdan

Bill Crosson, Maury Estes, Ashutosh Limaye, Dale Quattrochi, Doug Rickman, Carol Watts

NASA/MSFC/NSSTC

Judy Qualters, Pamela Meyer

Centers for Disease Control and Prevention

Amanda Niskar

Israel Center for Disease Control

Partners

Kaiser-Permanente Georgia
U.S. Environmental Protection Agency
Georgia Environmental Protection Division
Georgia Division of Public Health
Emory University
Georgia Institute of Technology

HELIX-Atlanta Overview

- HELIX-Atlanta was developed to support current and future state and local EPHT programs to implement data linking demonstration projects which could be part of the EPHT Network.
- HELIX-Atlanta is a pilot linking project in Atlanta for CDC to learn about the challenges the states will encounter.
- NASA/MSFC and the CDC are partners in linking environmental and health data to enhance public health surveillance.
- The use of NASA technology creates value – added geospatial products from existing environmental data sources to facilitate public health linkages.
- Proving the feasibility of the approach is the main objective

HELIX-Atlanta Challenges

- Sharing data between agencies with different missions and mindsets
- Protecting confidentiality of information
- Ensuring high quality geocoded data
- Ensuring appropriate spatial and temporal resolutions of environmental data
- Developing sound resources and methods for conducting data linkages and data analysis

HELIX-Atlanta Respiratory Health Team

RH Team Pilot Data Linkage Project:

Link environmental data related to ground-level $PM_{2.5}$ (NASA+EPA) with health data related to asthma

Goals:

6. **Produce and share information on methods useful for integrating and analyzing data on asthma and $PM_{2.5}$ for environmental public health surveillance.**
7. **Generate information and recommendations valuable to sustaining surveillance of asthma with $PM_{2.5}$ in the Metro-Atlanta area.**

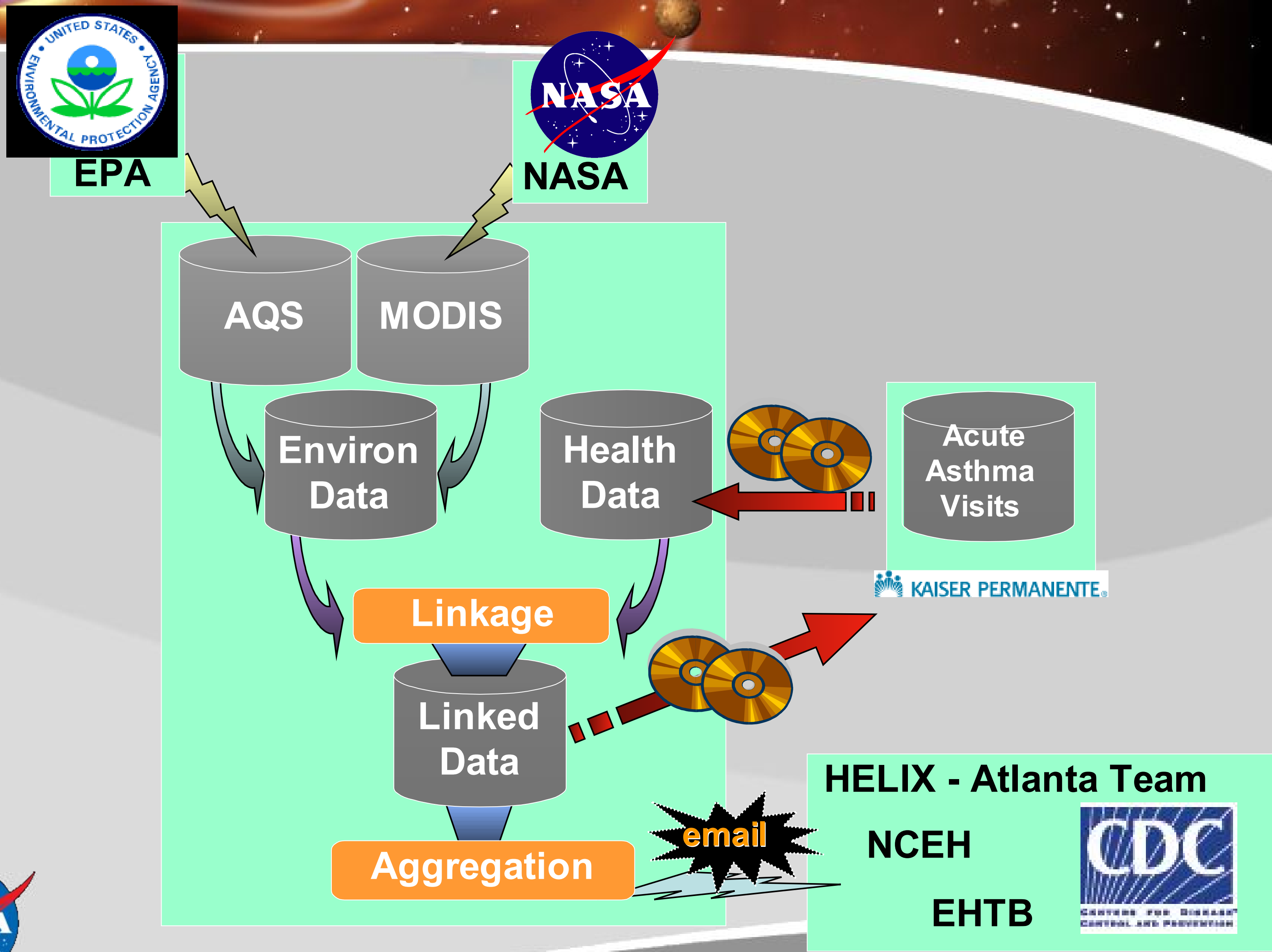
Environmental Hazard Measure: Daily $PM_{2.5}$

Asthma Measure: Daily acute asthma office visits to KP-GA Medical Facilities

Time period: 2001-2003

Linkage Domain: 5-county metropolitan Atlanta

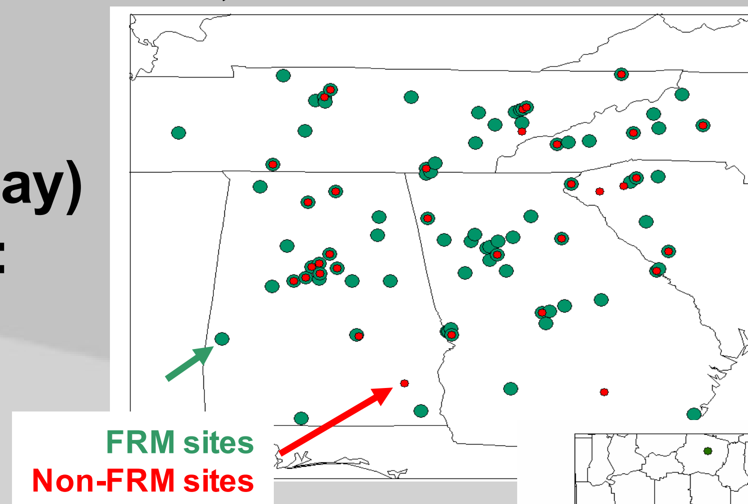
Data Linkage

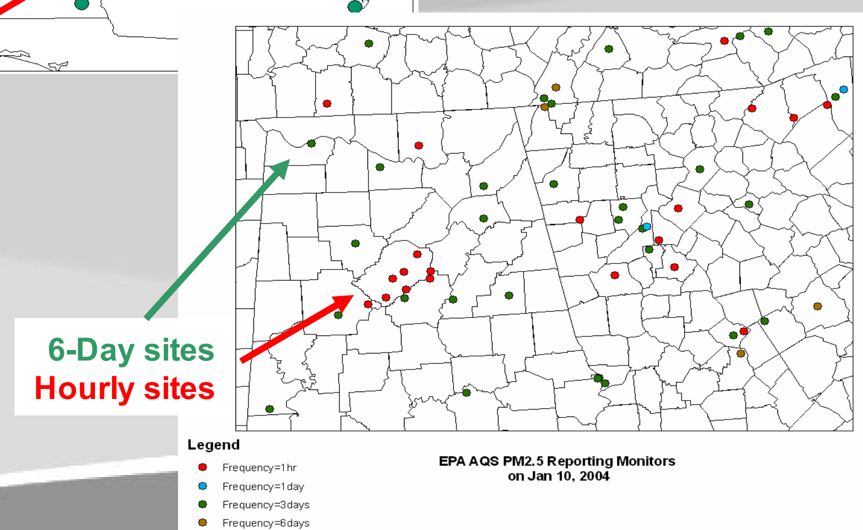


Sources of $PM_{2.5}$ data: EPA AQS

EPA Air Quality System (AQS) ground measurements

- National network of air pollution monitors
- Concentrated in urban areas, fewer monitors in rural areas
- Time intervals range from 1 hr to 6 days (daily meas. every 6th day)
- Three monitor types:
  - Federal Reference Method (FRM)
  - Continuous
  - Speciation
- FRM is EPA-accepted standard method; processing time 4-6 weeks

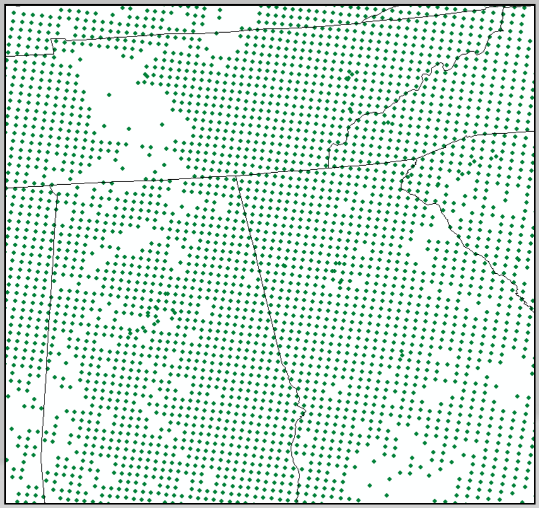




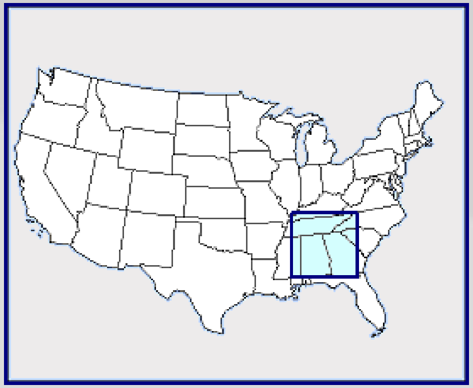
Sources of $PM_{2.5}$ data: MODIS

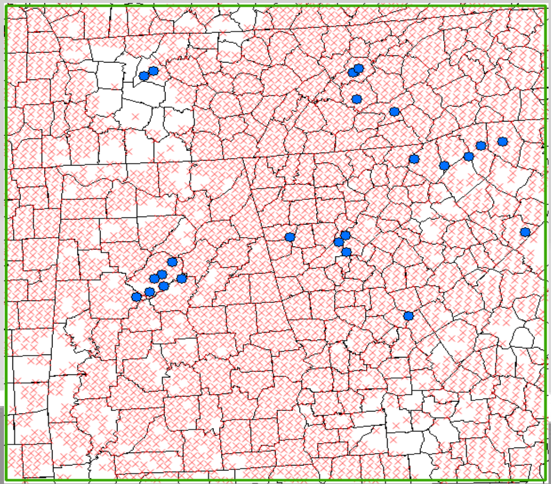
<u>MODIS Aerosol Optical Depth (AOD)</u>

- AOD is a measure of the total particulate in the atmosphere
- If atmosphere is well mixed, AOD is a good indicator of surface $PM_{2.5}$
- Enhanced Spatial Coverage
- Provided on a 10x10 km grid
- Available twice per day
(Terra ~10:30 AM, Aqua ~1:30 PM)
- Clear-sky coverage only
- Available since spring 2000



MODIS





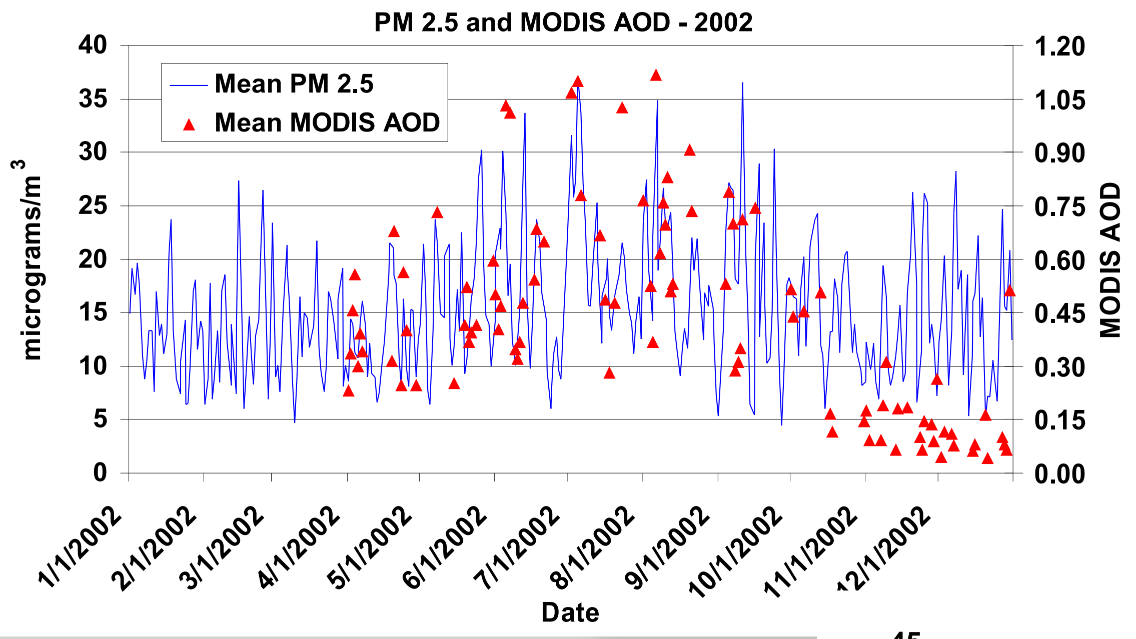
AQS

June 25, 2003

Estimating $PM_{2.5}$ from MODIS data

- For 2002-2003, obtain MODIS AOD and EPA AQS $PM_{2.5}$ data
- Extract AOD data for 5 AQS site locations
- Calculate daily averages from hourly AQS $PM_{2.5}$ data
- Using daily $PM_{2.5}$ averages from all 5 Atlanta AQS sites, determine statistical regression equations between $PM_{2.5}$ and MODIS AOD
- Apply regression equations to estimate $PM_{2.5}$ for each 10 km grid cell across region

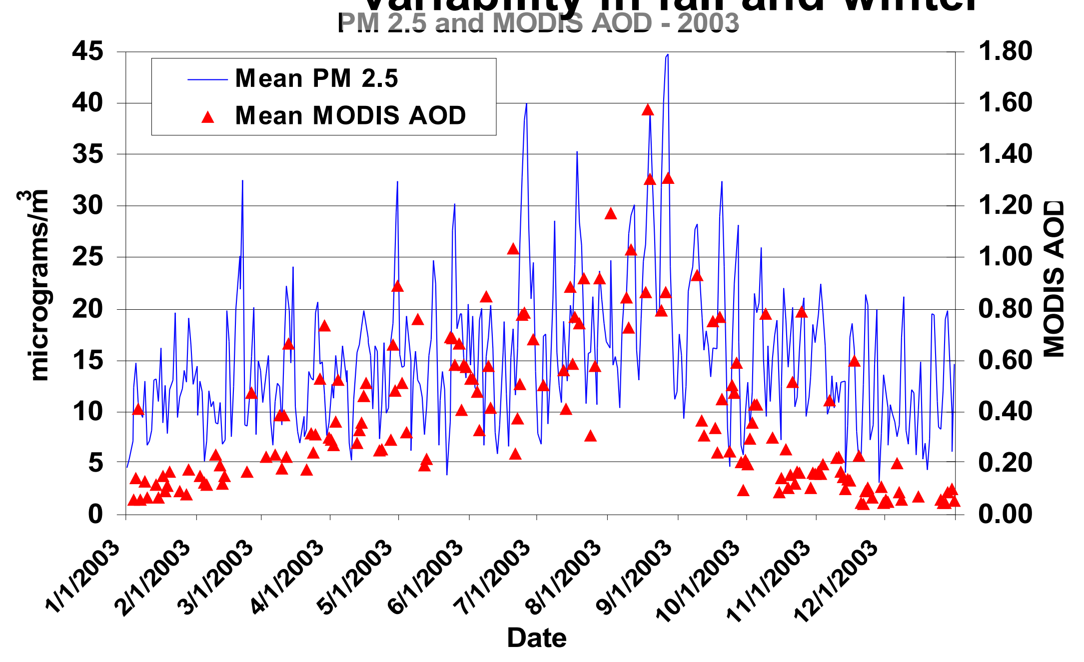
MODIS AOD - $PM_{2.5}$ Relationship



- Daily 5-site means of observed $PM_{2.5}$ and MODIS AOD
- MODIS data not available every day due to cloud cover
- MODIS AOD follows seasonal patterns of $PM_{2.5}$ but not the day-to-day variability in fall and winter

2002

2003



PM 2.5 – MODIS AOD Correlations

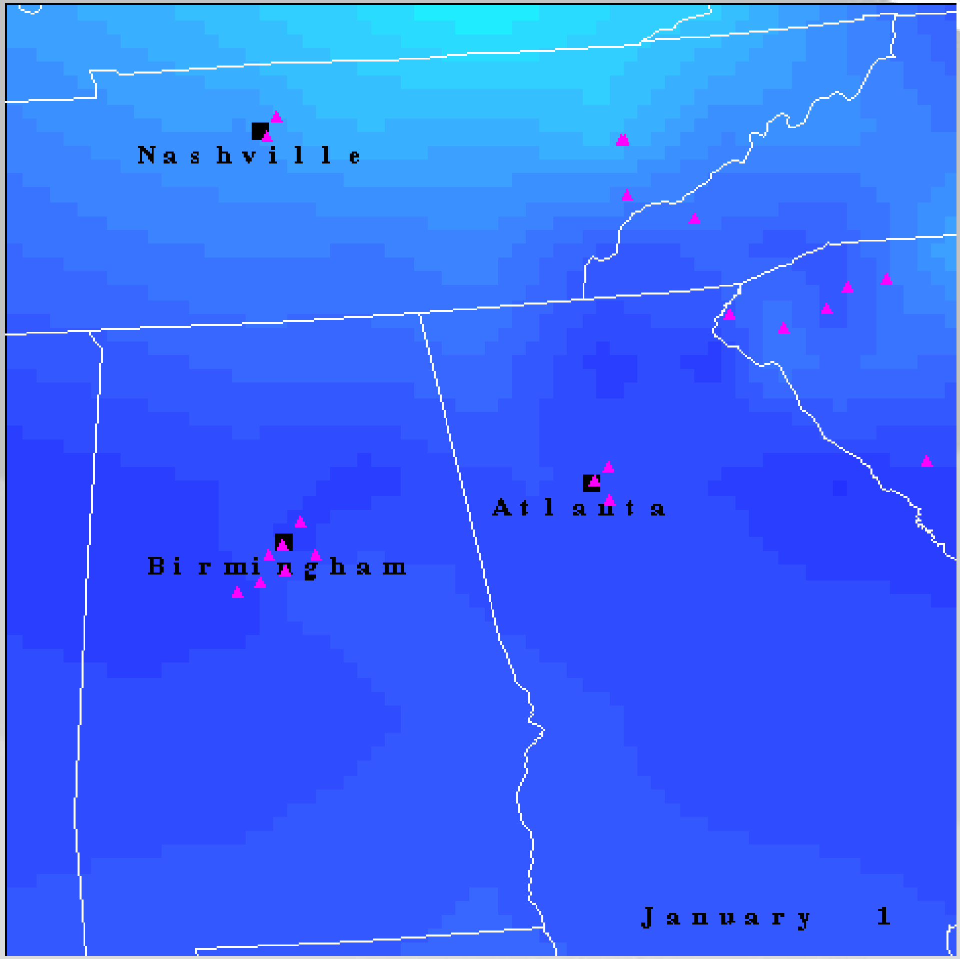
April - September

	MODIS-Terra	MODIS-Aqua
2000 -->	0.579	
2001 -->	0.643	
2002 -->	0.559	0.401
2003 -->	0.661	0.727

- **Correlations between $PM_{2.5}$ and MODIS AOD are generally high (> 0.55) for the warm season.**
- **The lower correlation for MODIS-Aqua in 2002 is for July-September only.**

PM2.5 Exposure Assessment- Spatial Surfacing

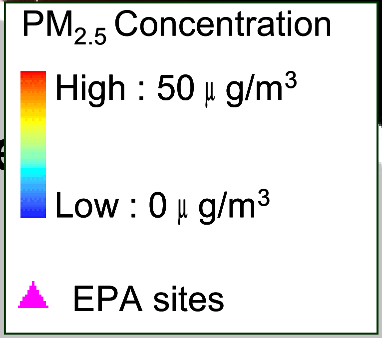
- 1st degree recursive B-spline in x- and y-directions
- Inverse Distance Weighted (IDW)
- Daily surfaces created on a 10x10 km grid
- Variable number of measurements available each day

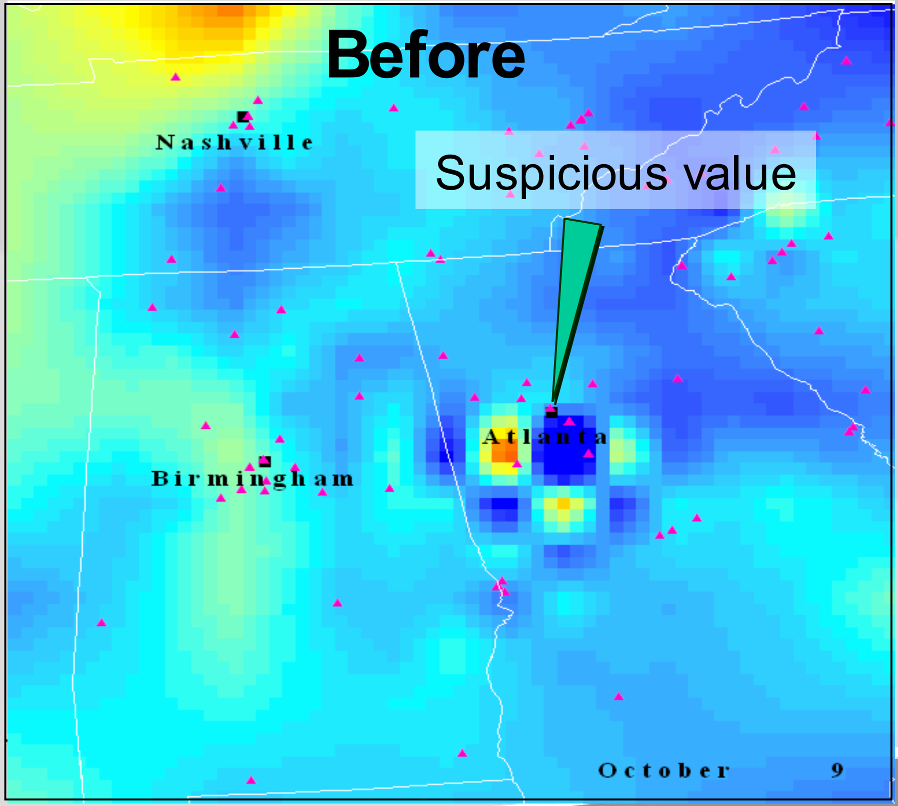


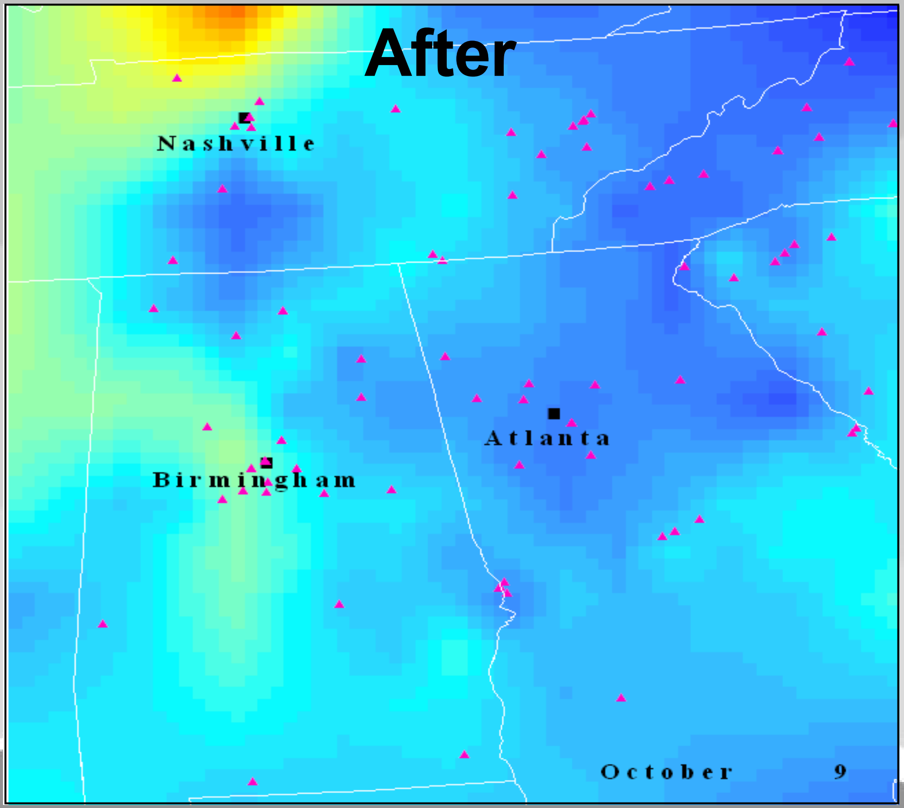
$PM_{2.5}$ Concentration
High : 50 µ g/m^3
Low : 0 µ g/m^3
EPA sites

Quality Control Procedure for AQS $PM_{2.5}$ data

- Eliminates anomalous measurements based on a non-parametric rank-order spatial analysis
- Applied to all daily AQS $PM_{2.5}$ measurements before spatial surfaces are built



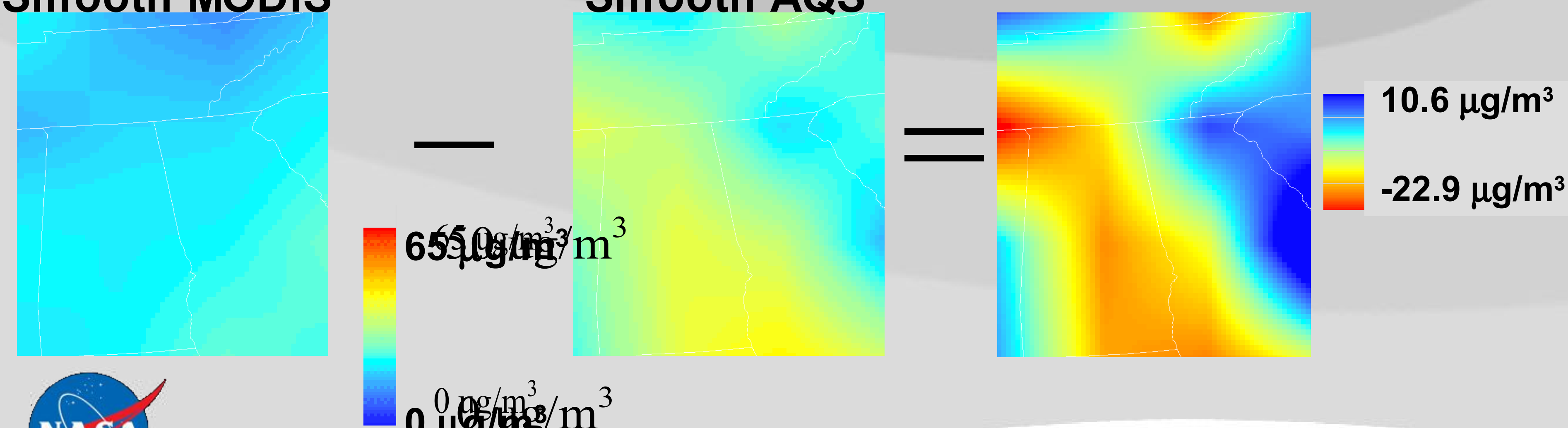




MODIS $PM_{2.5}$ Bias Adjustment

- Assumption: AQS measurements are unbiased relative to the local mean, but MODIS $PM_{2.5}$ estimates may have biases.
- Procedure:
  1. Use a two-step B-spline algorithm to create highly smoothed versions of the MODIS and AQS $PM_{2.5}$ daily surface
  2. Compute the 'Bias' as the difference between the smoothed fields
  3. Subtract the bias from the MODIS $PM_{2.5}$ daily surface to give the 'bias-corrected' MODIS daily surface

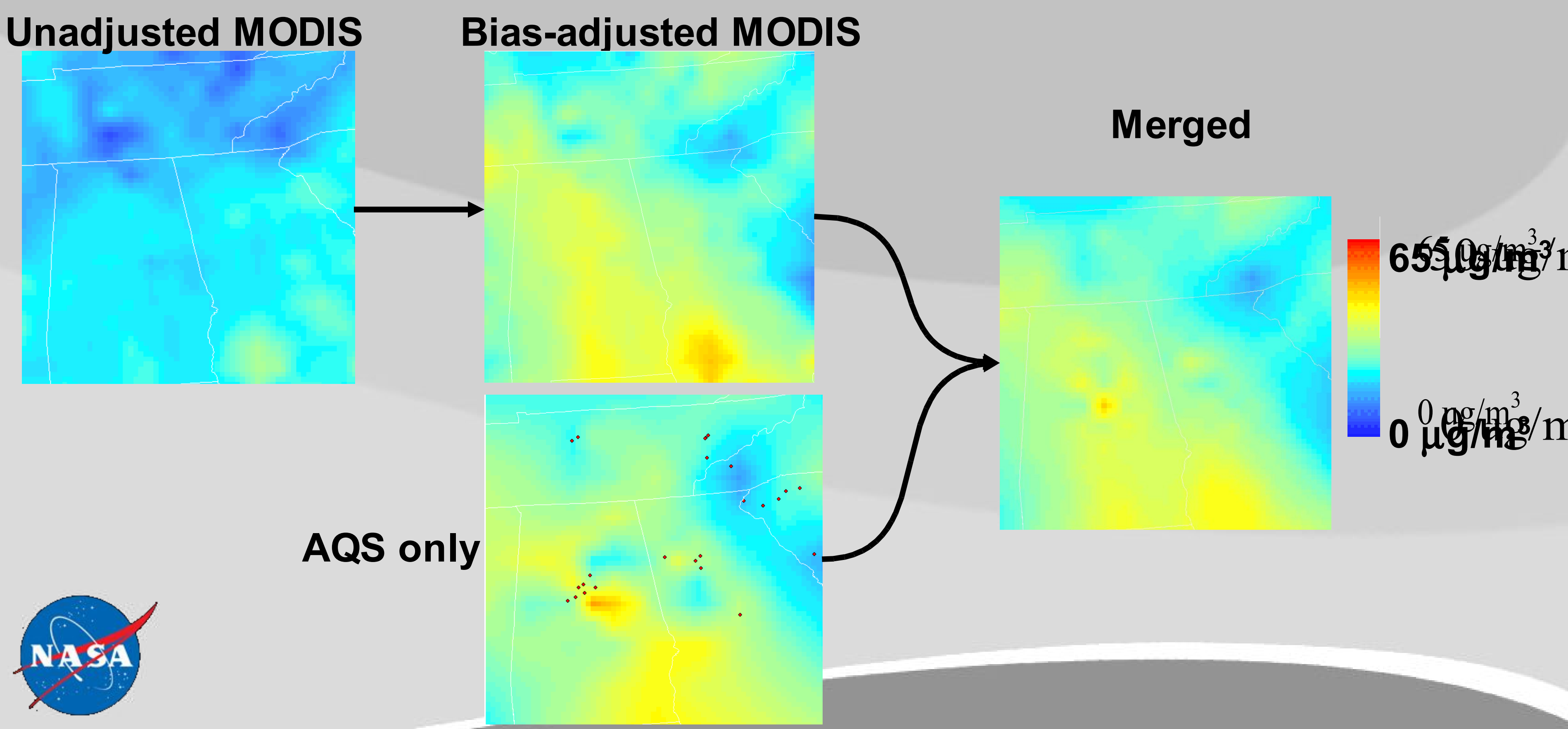
Smooth MODIS — Smooth AQS = MODIS Bias



Merging MODIS and AQS $PM_{2.5}$ Data

- MODIS and AQS data have been merged to produce final $PM_{2.5}$ surfaces.

B-Spline Surfacing

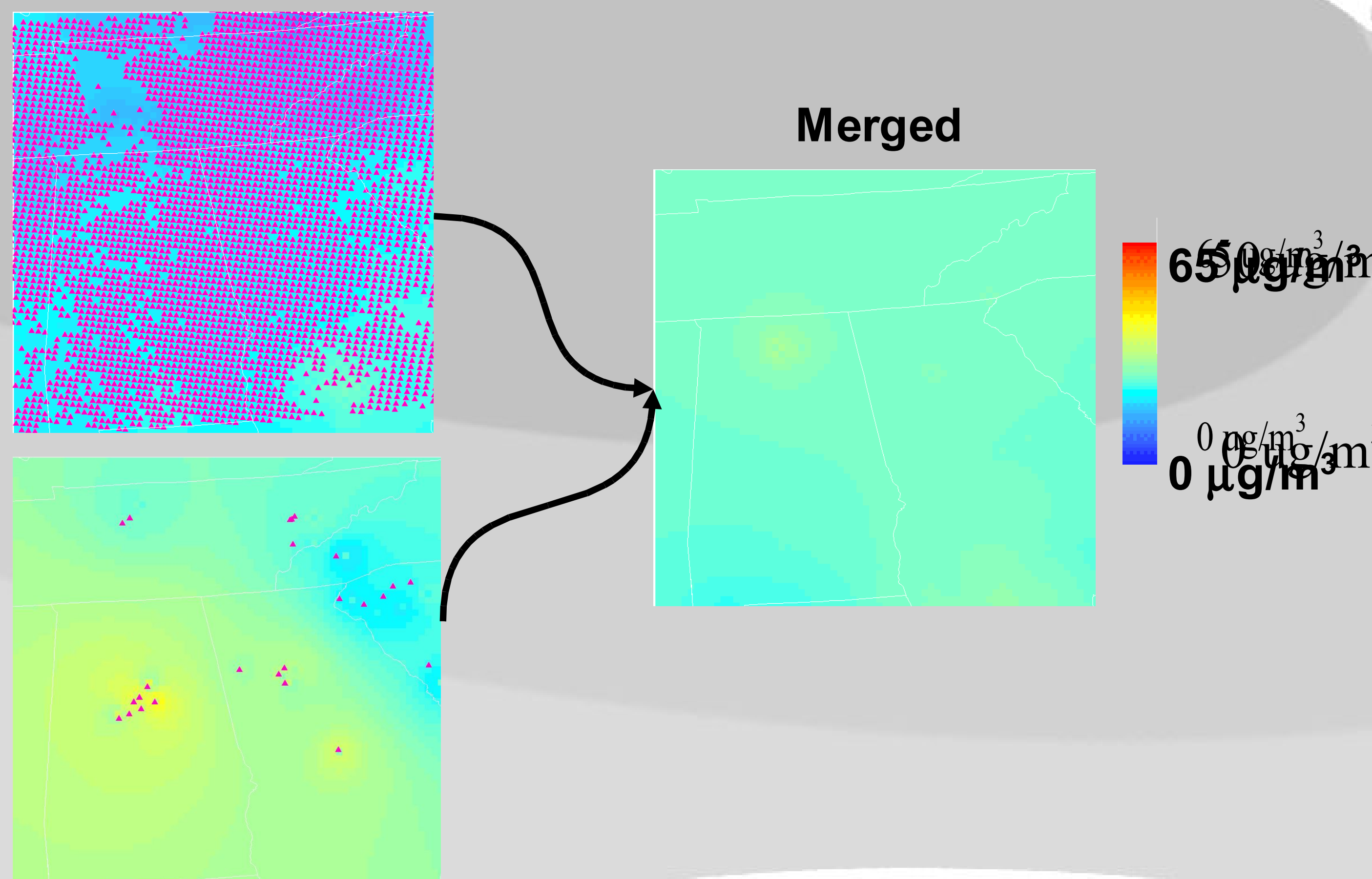


Merging MODIS and AQS $PM_{2.5}$ Data

IDW Surfacing

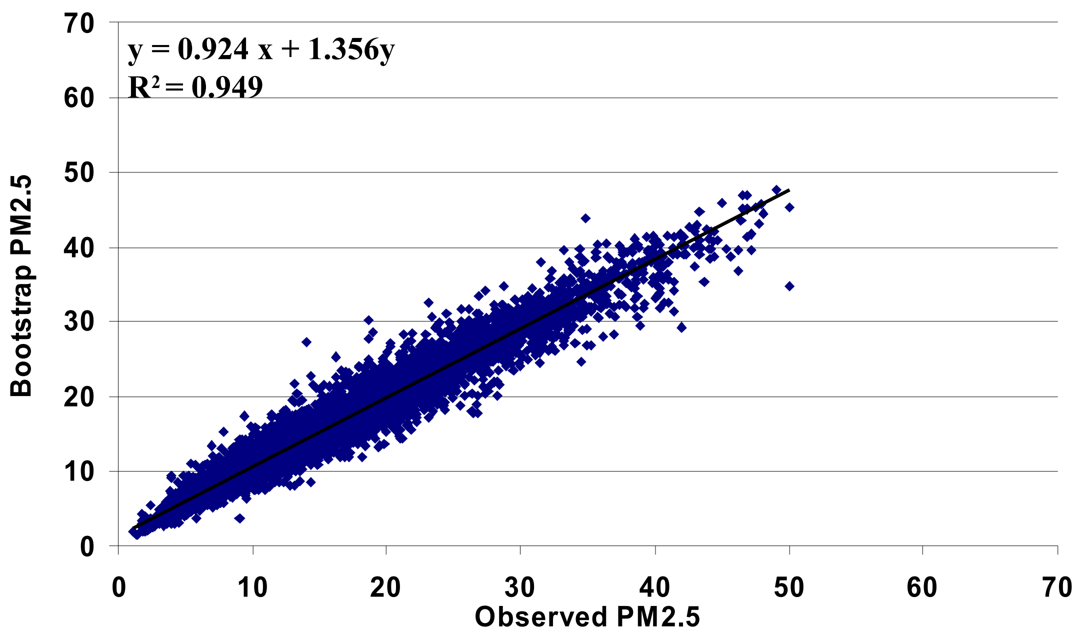
MODIS Only

AQS only

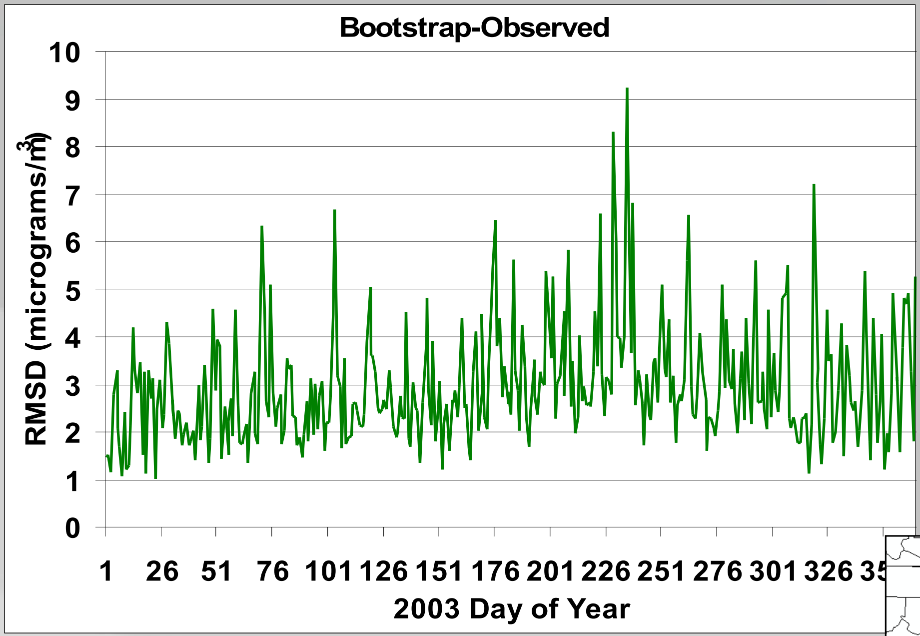


Cross-Validation

- a.k.a. ‘bootstrapping’ or ‘omit-one’ analysis
- Objective: Estimate errors associated with daily spatial surfaces
- Procedure:
  1. Omitting one observation, create surface using N-1 observations
  2. Compare value of surface at location of omitted observation with the observed value
  3. Repeat for all observations
  5. Calculate error statistics by day or site



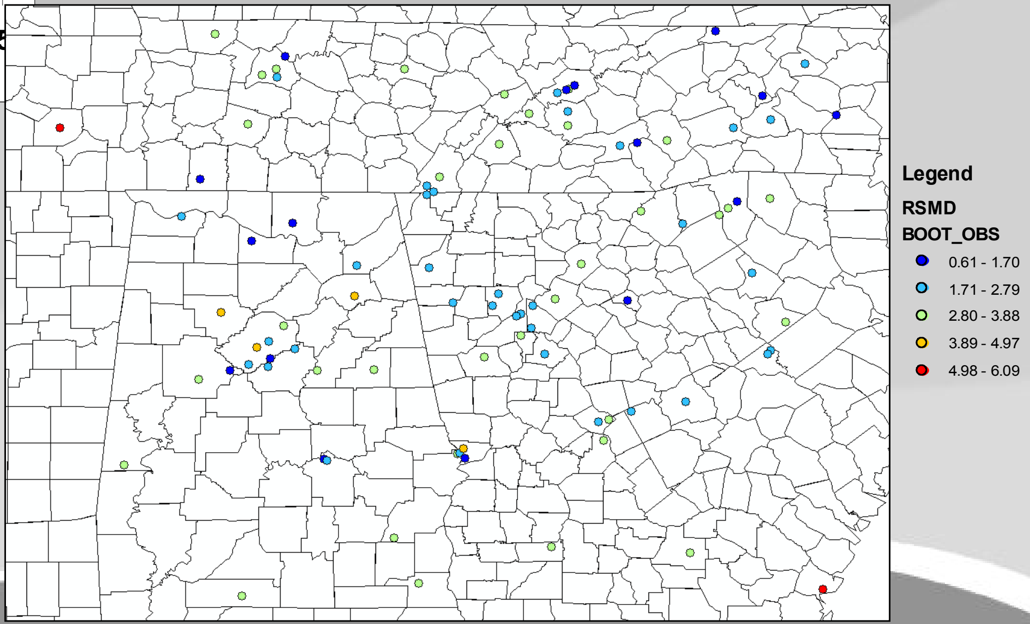
Cross-Validation Error Statistics



Time Series

RMSD = 2.7 $\mu g/m^3$

RMSD by Site

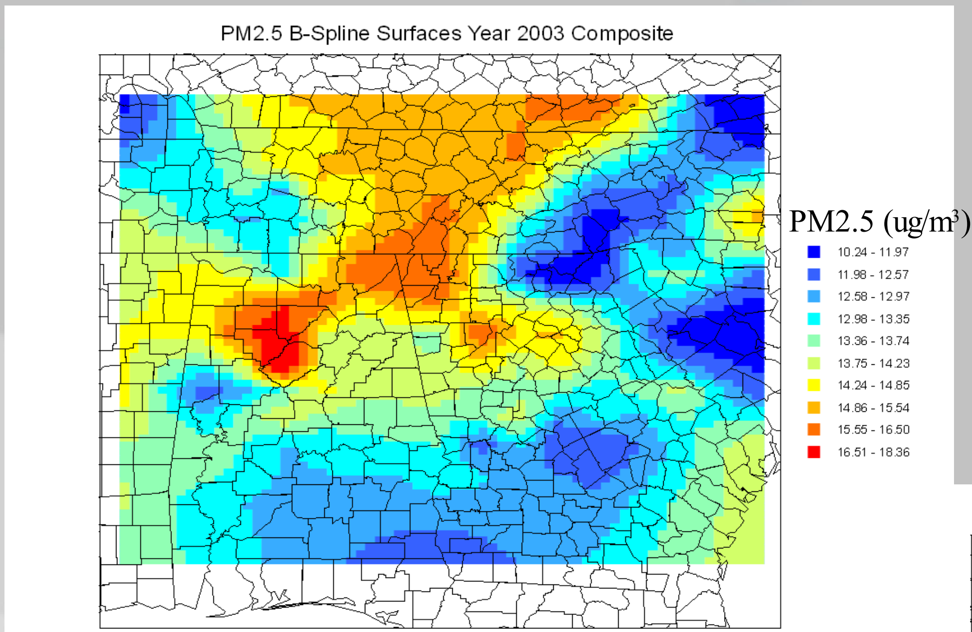


Surfacing Methods Comparison

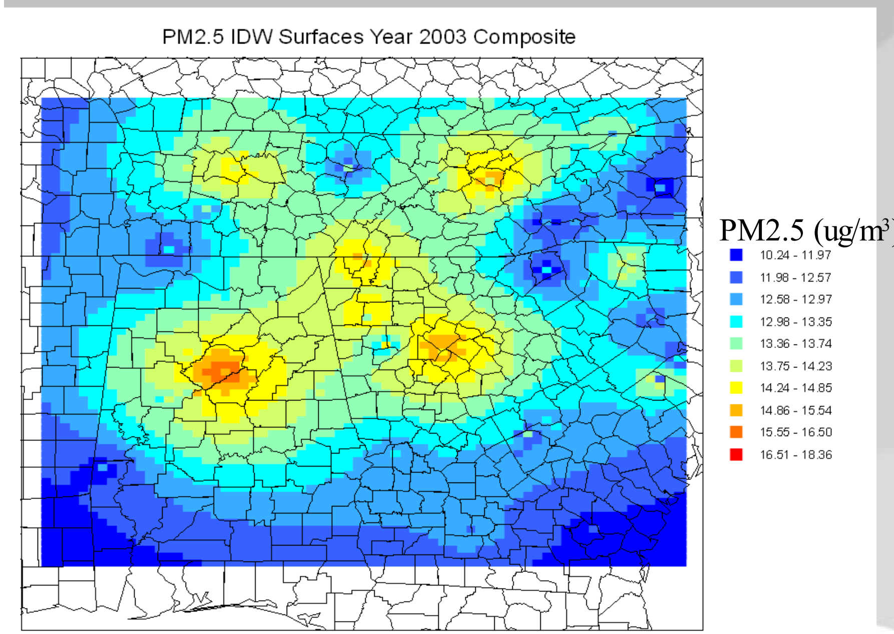
Surfacing Technique and Data Source	RMSD (All Days)	RMSD (Warm Season (Days 91-273))
Bspline, AQS only, no QC	3.302	3.556
Bspline, AQS only, with QC	2.927	3.164
IDW, AQS only	2.450	2.686
B-Spline, merged AQS/MODIS	N/A	2.756
IDW, merged AQS/MODIS	N/A	1.613

Surfacing Technique and Data Source	Improvement
Bspline: QC vs. No QC	12 %
Bspline: AQS only vs. merged AQS/MODIS	16 %
IDW: AQS only vs. merged AQS/MODIS	40 %

Annual Composite Surfaces



B-Spline



IDW

Linkage of Environmental and Health Data

Health Data Set

Members

LON	LAT	ID	AGE	GENDER	YEAR/MO
-84.207	99.200	1	Child	M	200301
-84.802	99.359	2	Adult	M	200301
-83.798	99.993	4	Child	F	200301

Acute asthma office visits

ID	AGE	LON	LAT	GENDER	DATE
1811	Child	-84.179	99.118	F	1/1/2003
54767	Adult	-84.625	99.802	F	1/1/2003
84580	Adult	-84.679	99.691	F	1/1/2003

*Simulated Data Set. F=female, M=male, A=adult, C=child.

Linkage of Environmental and Health Data

Data Linkage Outputs

Visit counts by grid cell

Date	Cell	PM2.5	FC	MC	FA	MA
200301	1	21.74	1	0	2	0
200301	2	12.79	0	0	0	0
200301	3	12.21	0	1	0	1

$PM_{2.5}$ for each visit

Date	ID	Member	Lat/Lon		Cell	Cell Lat/Lon		County	State	Gender	Age	PM2.5
1	1	1811	99.572	-84.251	1944	99.552	-84.284	Coweta	GA	F	Child	21.74
1	2	15299	99.063	-83.860	1608	99.104	-83.806	Upson	GA	F	Child	12.79
1	2	15879	99.727	-84.369	2079	99.731	-84.403	Fulton	GA	M	Child	12.21

*Simulated Data Set. F=female, M=male, A=adult, C=child.

Successes

- Proven the feasibility of linking environmental data (MODIS $PM_{2.5}$ estimates and AQS) with health data (asthma)
- Developed algorithms for QC, bias removal, merging MODIS and AQS $PM_{2.5}$ data, and others…
- Negotiated a Business Associate Agreement with a health care provider to enable sharing of Protected Health Information

Team Members and Acknowledgements

Member's Name, Affiliation

- *(Co-Chair)* Kafayat Adeniyi, Centers for Disease Control and Prevention,
- *(Co-Chair)* Solomon Pollard, Environmental Protection Agency (EPA), Region 4
- Mohammad Z. Al-Hamdan, National Aeronautics and Space Administration
- Rob Blake, DeKalb County Board of Health
- David Blaney, Georgia Division of Public Health
- Bill Crosson, National Aeronautics and Space Administration
- Kristen Mertz, Georgia Division of Public Health
- Amanda Sue Niskar, Centers for Disease Control and Prevention
- Dale Quattrochi, National Aeronautics and Space Administration
- Amber Sinclair, Kaiser Permanente
- Allison Stock, Centers for Disease Control and Prevention
- Denis Tolsma, Kaiser Permanente
- Linda Thomas, Environmental Protection Agency, Region 4
- Ntale Kajumba, Environmental Protection Agency, Region 4
- Carolyn Williams, Georgia Division of Public Health

Acknowledgments

- Leslie Fierro, Centers for Disease Control and Prevention
- Gabriel Rainisch, Centers for Disease Control and Prevention
- Emily Hansen
- HELIX-Atlanta Partners

Thanks!

Surfacing Technique (IDW)

Inverse Distance Weighted Interpolation (IDW)

• IDW determines cell values using a weighted combination of a set of observational points. The weight is a function of inverse distance. The further the point from the cell, the less effect it will have on the interpolated value of that cell.

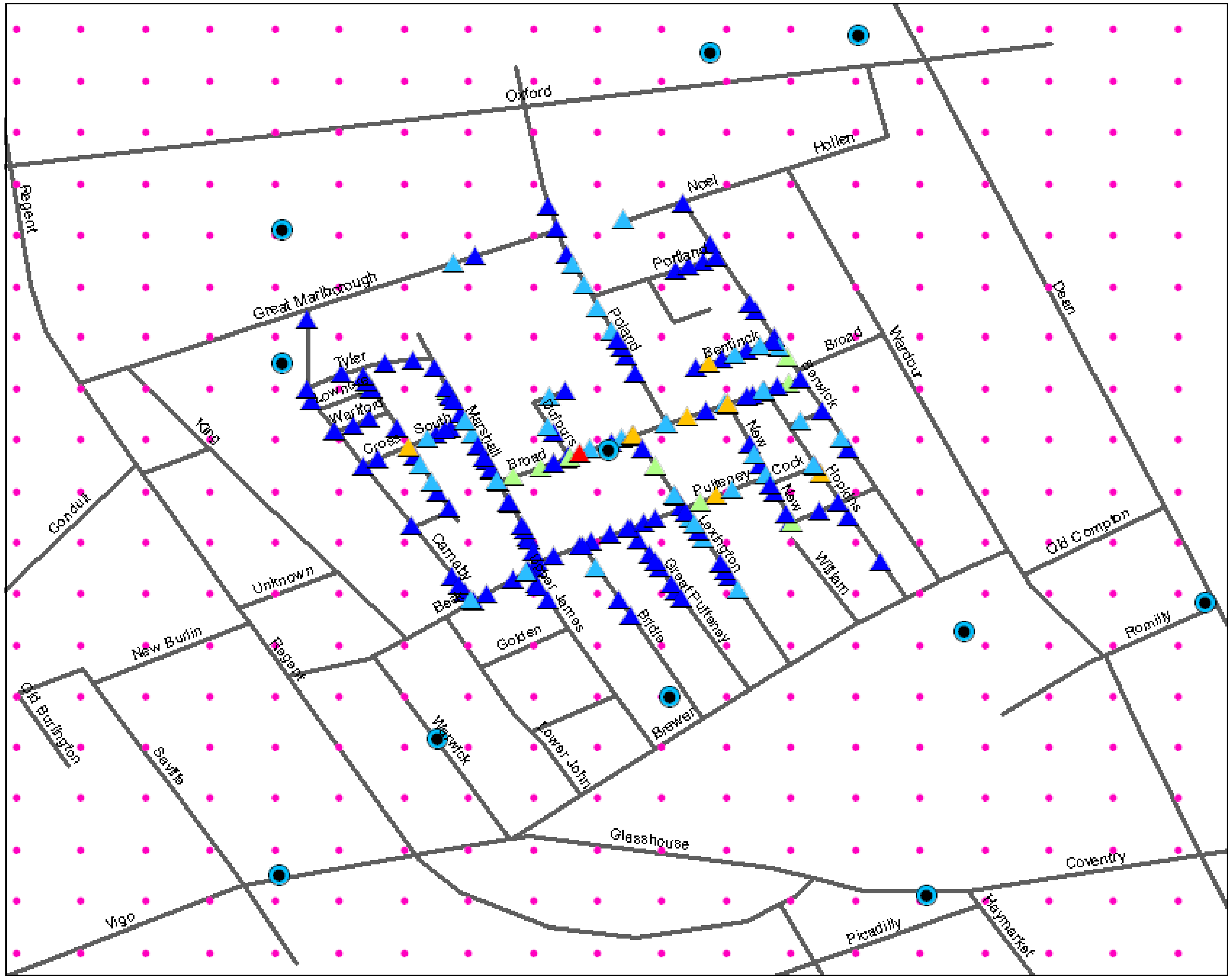
Strengths:
•Simplicity of underlying principle
•The speed in calculation for small data sets

Weaknesses:
•Equal assigned weight to each of the data points even if it is in a cluster
•Maxima and Minima in the interpolated surface can only occur at data points
•Creates high frequency artifacts
•Creates artifacts with wavelengths comparable to sample spacing
•It is an interpolating logic
•Spurious values have effect throughout the entire surface
•The speed of calculation for large data sets

Surfacing Technique (B-Spline)

- A recursive B-Spline of degree 1st degree x- and y-directions
    - Flexibility
    - Robust treatment of noise and spurious values
    - Resistant to artifacts and introduction of spurious frequencies
    - Handles data density and distribution issues better than most algorithms
    - Result closely approximates what you do by hand
    - Limitation – It does not handle discontinuities in the assumed surface w/o advanced programming logic
- There are a few simple controls, which give much flexibility

Cholera Deaths Soho, London August-September, 1854



Legend

- Streets
- Wells
- Grid

Cholera Deaths Per Residence

COUNT

- 1 - 2
- 3 - 4
- 5 - 6
- 7 - 10
- 11 - 18

Integration Radius = 55m
Grid Spacing = 40m

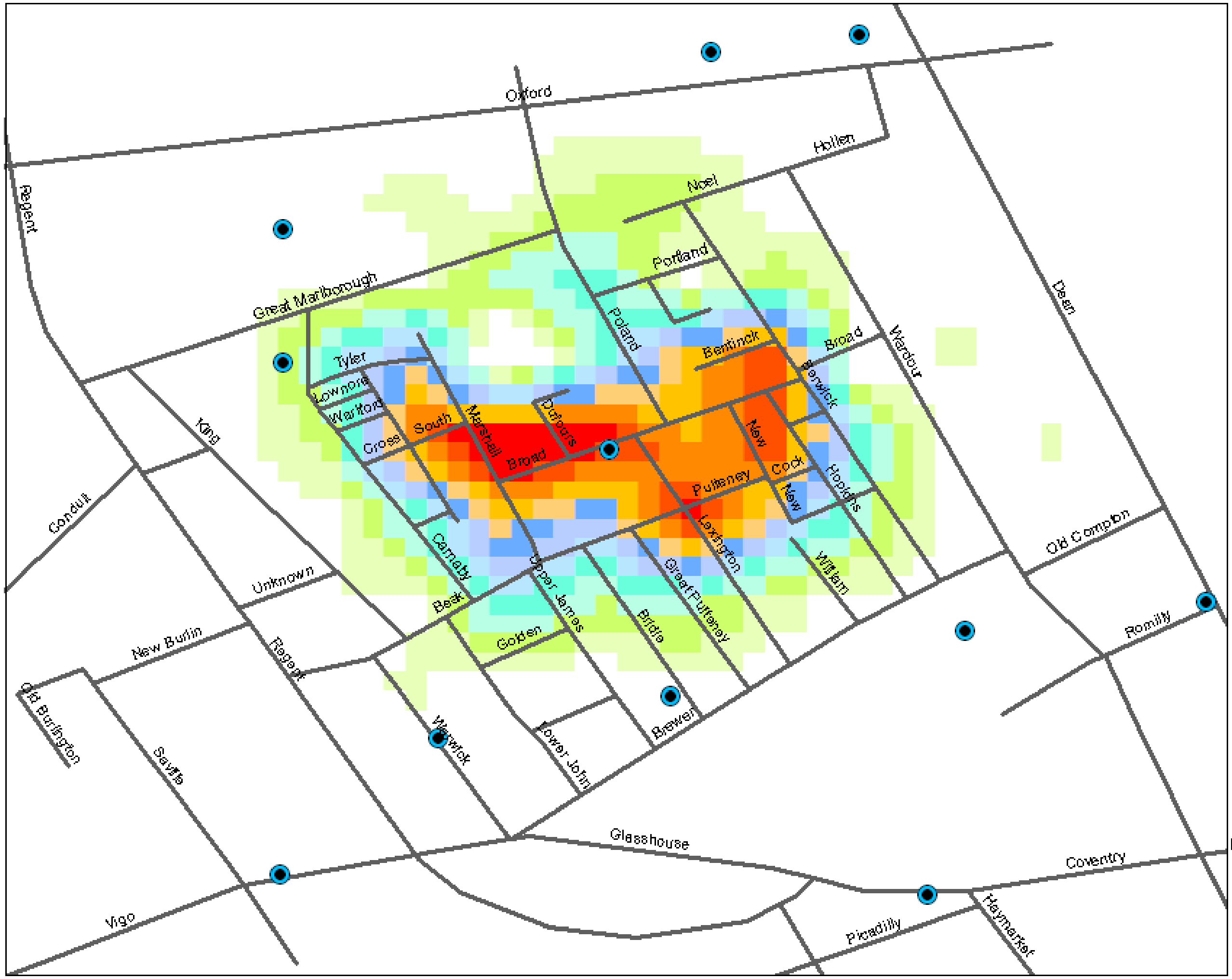
N

Meters
0 50 100 200

12/15/05 10:35am
Dr. Mohammad Al-Hamdan
USRA at NASA/MSFC
HELIX-Atlanta Project
mohammad.alhamdan@nasa.gov

*Original data were published by C.F. Cheffins, Lith, Southhampton Buildings, London, England, 1854 in Snow, John. On the Mode of Communication of Cholera, 2nd Ed, John Churchill, New Burlington Street, London, England, 1855.
**Digital Data of Streets, Wells, and Deaths Residences which were used to creat this surface were downloaded from the UCLA Department of Epidemiology Website at http://www.ph.ucla.edu/epi/snow.html.

Cholera Deaths Soho, London August-September, 1854



Legend

- Streets
- Wells

Deaths Per Unit Area

COUNT

- 0 - 1
- 1 - 5
- 5 - 10
- 10 - 15
- 15 - 20
- 20 - 25
- 25 - 30
- 30 - 35
- 35 - 40
- 40 - 45
- 45 - 50
- 50 - 55
- 55 - 63

Integration Radius = 55m
Grid Spacing = 40m

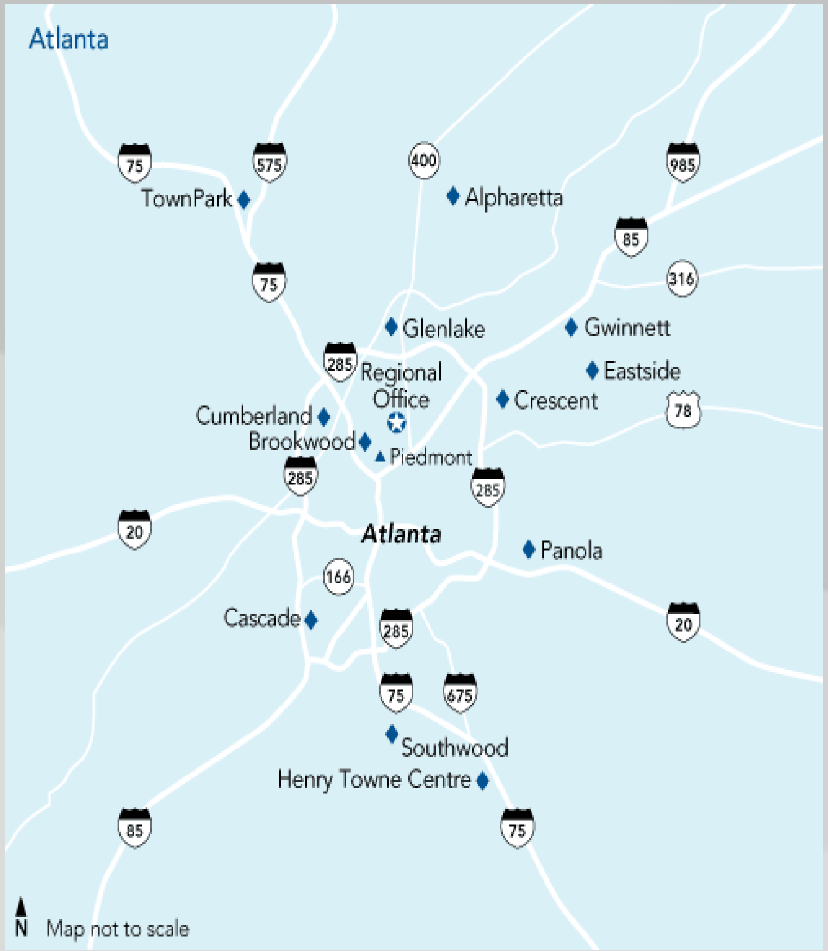
N

Meters
0 50 100 200

12/15/05 10:35am
Dr. Mohammad Al-Hamdan
USRA at NASA/MSFC
HELIX-Atlanta Project
mohammad.alhamdan@nasa.gov

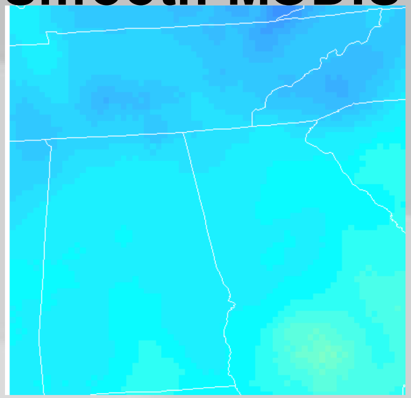
*Original data were published by C.F. Cheffins, Lith, Southhampton Buildings, London, England, 1854 in Snow, John. On the Mode of Communication of Cholera, 2nd Ed, John Churchill, New Burlington Street, London, England, 1855.
**Digital Data of Streets, Wells, and Deaths Residences which were used to creat this surface were downloaded from the UCLA Department of Epidemiology Website at http://www.ph.ucla.edu/epi/snow.html.

Kaiser Permanente Georgia (KP-GA)



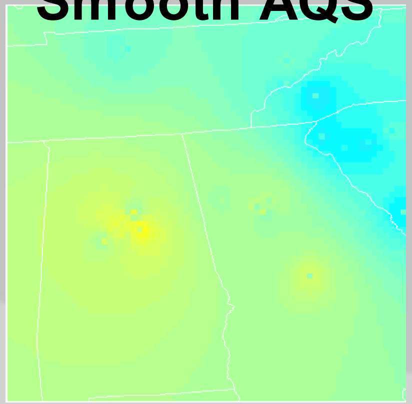
- 13 Kaiser facilities
- 264,708 Members (7/04)
- 90% Group Model
- Diverse Membership
- Acute Visit Access
- 20 county metro area coverage
- Contract hospitals for emergency care
- Mean 2.7 primary care visits per member/year
- Mean 17.6 acute child asthma visits/day
- Mean 11.8 acute adult asthma visits/day

Smooth MODIS



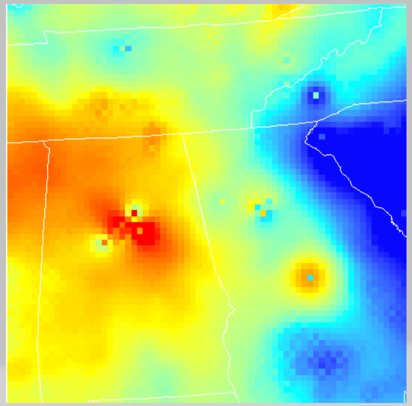
−

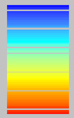
Smooth AQS



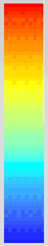
=

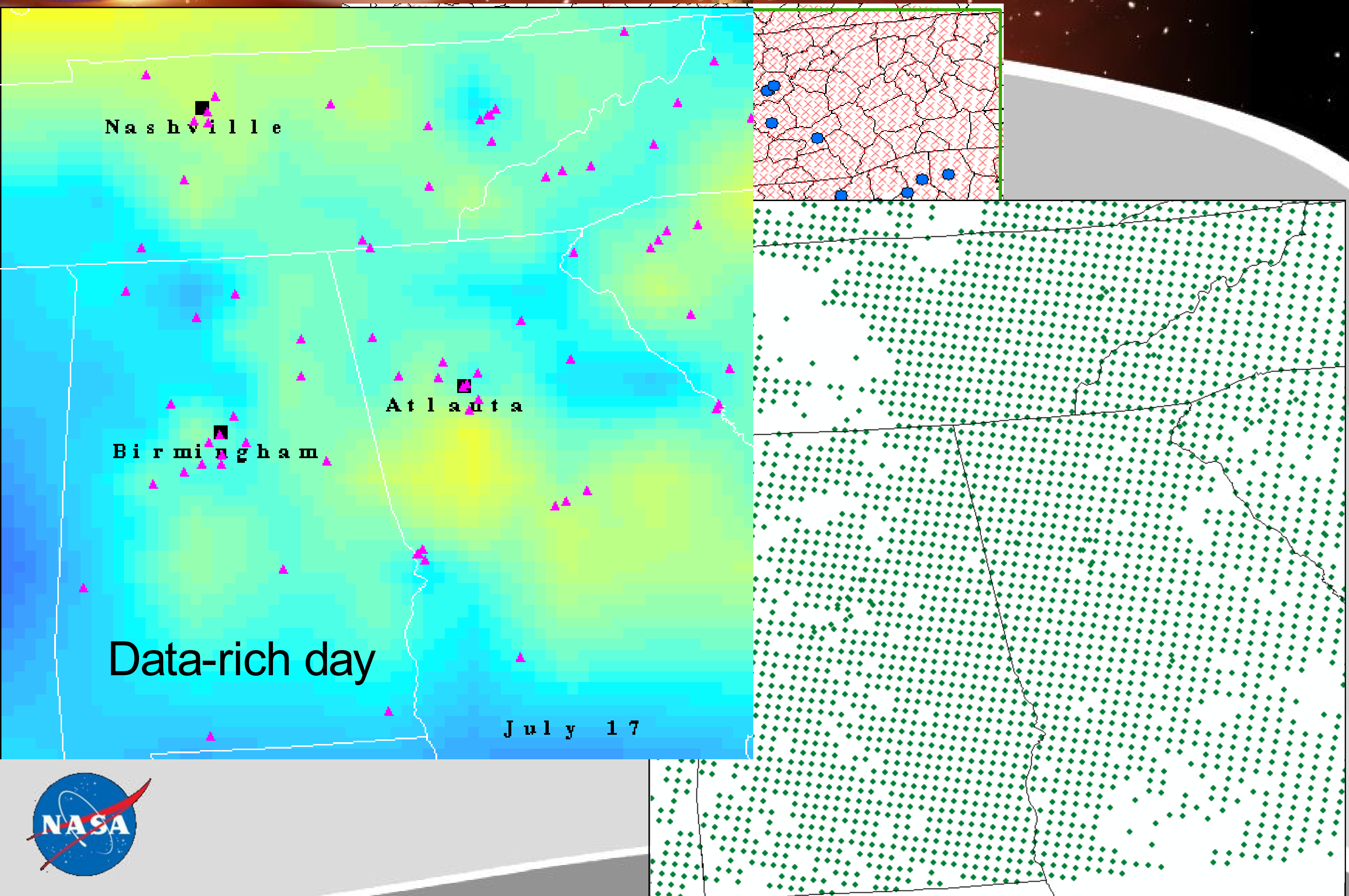
MODIS Bias



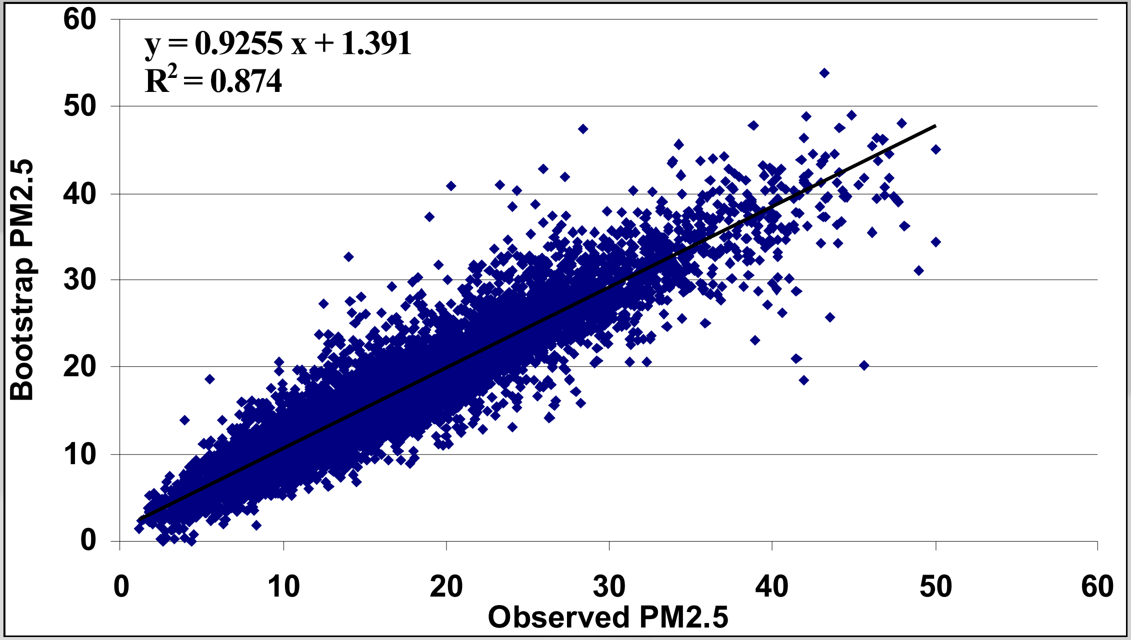


IDW





Nashville
Atlanta
Birmingham
Data-rich day
July 17


y = 0.9255 x + 1.391
R2 = 0.874
Bootstrap PM2.5
Observed PM2.5
0
10
20
30
40
50
60

Remaining Challenges

- Build computer infrastructure to enable public health surveillance
- Identify and develop environment data sources from NASA or elsewhere that are better suited for public health surveillance
- Coordinate with state and local agencies to develop public health surveillance networks in their locales